DOCUMENT RESUME

ED 266 728 HE 019 098

AUTHOR Whitman, Neal A.; Cockayne, Thomas W.
TITLE Evaluating Medical School Courses: A User-Centered Handbook.
INSTITUTION Utah Univ., Salt Lake City. School of Medicine.
PUB DATE 84
NOTE 40p.
AVAILABLE FROM University of Utah School of Medicine, Department of Family & Community Medicine, 50 North Medical Drive, Salt Lake City, UT 84132 ($12.00).
PUB TYPE Guides - Classroom Use - Guides (For Teachers) (052) -- Tests/Evaluation Instruments (160)

EDRS PRICE MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS Check Lists; Committees; *Course Evaluation; Department Heads; Ethics; *Evaluation Criteria; Evaluation Methods; *Faculty Evaluation; Higher Education; *Medical Education; *Medical School Faculty; *Needs Assessment; Peer Evaluation; Profiles; Rating Scales; Student Evaluation of Teacher Performance

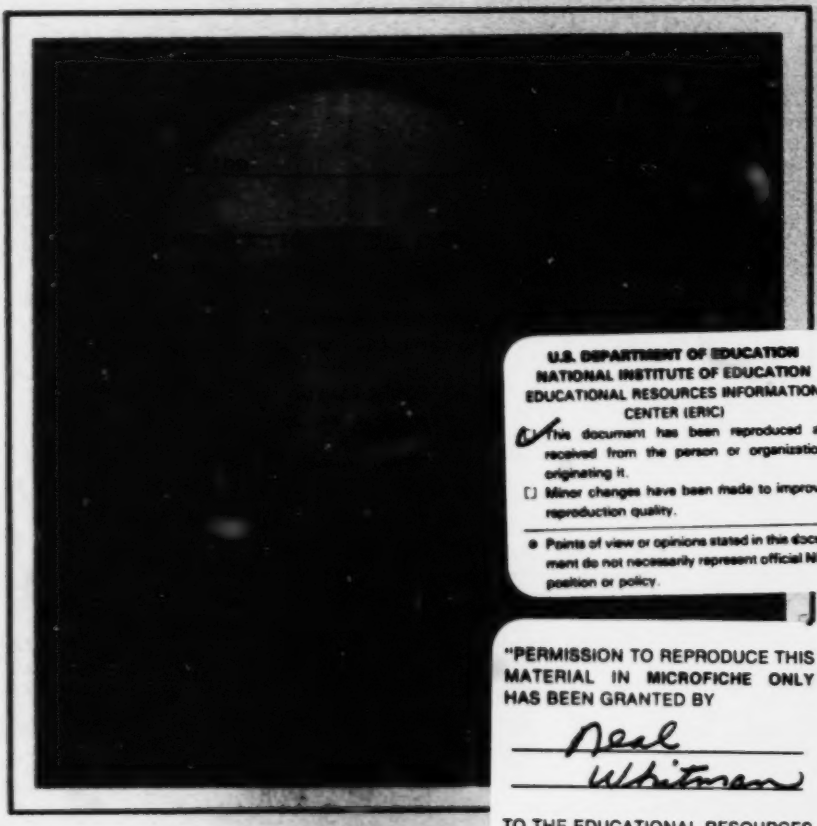
ABSTRACT

A handbook to help faculty plan and implement the evaluation of medical school courses is presented. Seven steps necessary to planning a course evaluation are specified: identify users and needs, ask specific questions, indicate how questions will be answered, select and/or design instruments, consider ethical issues, collect and analyze data, and report results. Questions to ask about a course might be based on perceptions of students, colleagues, and the teacher of the course; the evaluations of educational specialists; and learning outcomes. To help implement course evaluations, four major users of medical school course evaluations need to be considered: course director and teachers, department head, school curriculum committee, and promotions committee. To assess the course evaluation, practitioners should consider the usefulness and practicality of the evaluations, technical adequacy, and ethics. A reference list with 31 items is included. Appendices include: a form for peer review of undergraduate teaching based on dossier materials; rating scales for student evaluation of the lecture and course instructor; a list of possible items for a teacher, including eight products of good teaching (e.g., test scores), material from oneself (e.g., list of course materials prepared for students), and information from students, colleagues, and others. (SW)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

EVALUATING MEDICAL SCHOOL COURSES: A USER-CENTERED HANDBOOK

ED266728



U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

This document has been reproduced as received from the person or organization originating it.

[ ] Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY

Neal Whitman

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

HE 019 098

Neal A. Whitman, Ed. D. Department of Family and Community Medicine

Thomas W. Cockayne, Ph.D. Office of Medical Education

University of Utah School of Medicine

This handbook is dedicated to Dr. G. Richard Lee and Dr. Anne Osborn, who as Dean of the Medical School and as Assistant Dean for Medical Education, were strong supporters of our efforts to improve and evaluate teaching at the University of Utah School of Medicine.

Neal A. Whitman directs educational development in the Department of Family and Community Medicine. He has worked as a teacher, administrator, and consultant in medical education since 1971. Dr. Whitman earned a Master of Arts degree in higher education from the University of Michigan, a Master of Public Administration in health planning from New York University, and a Doctor of Education degree from Teachers College, Columbia University.

Thomas W. Cockayne has been Director of the Office of Medical Education at the University of Utah School of Medicine since 1974. During that time he has consulted with medical school faculty concerning curriculum development and evaluation. Dr. Cockayne obtained a Ph.D. in Educational Psychology in 1974 from the University of Utah.

Copyright © 1984 by the
University of Utah School of Medicine
Salt Lake City, Utah 84132

All Rights Reserved

Printed in the United States of America

TABLE OF CONTENTS

INTRODUCTION ..... page 1

PART I. PLANNING COURSE EVALUATIONS ..... page 1

A. Identify users and needs.
B. Ask specific questions.
C. Indicate how questions will be answered.
D. Select and/or design instruments.
E. Consider ethical issues.
F. Collect and analyze data.
G. Report Results.

PART II. USER-CENTERED EVALUATION ..... page 17

A. Evaluation for Course Directors and Teachers.
B. Evaluation for Department Heads.
C. Evaluation for Curriculum Committee.
D. Evaluation for Promotion Committee.

PART III. META EVALUATION ..... page 23

REFERENCES 25

APPENDICES: A ..... page 27
B ..... page 29
C ..... page 32
D ..... page 33
E ..... page 34
F ..... page 35

Introduction

The purpose of this handbook is to help faculty *plan* and *implement* the evaluation of medical school courses. Although evaluating courses in a medical school is not very different from evaluating courses in other types of higher education, the authors believe that it is useful to offer a handbook tailored to the medical school setting. Certainly, feedback from medical school faculty has encouraged us to believe so. For readers not in the medical setting, we hope that you will apply the principles to your courses, as well.

To help you *plan* course evaluations, a series of steps will be described. By way of preview, the first step is to identify the users and their needs. The Joint Committee on Standards for Evaluation (which will be discussed in more detail later) warns that if evaluation users and their needs are not clearly identified in advance,

> . . . the evaluation may be a misguided, academic exercise whose results are ignored, criticized, or resisted because they do not address anyone's particular questions. On the other hand, an evaluation planned and conducted to meet the information needs of identified audiences is more likely to receive a positive response (1981, p. 21).

By following the recommended steps and planning the course evaluation, we wish to help you avoid wasting time and energy.

To help you *implement* course evaluation, a section of the handbook is devoted to each of the four major users of medical school course evaluations: the course director and teachers; the department head; the school curriculum committee; and the promotions committee.

In addition to helping you plan and evaluate course evaluations, we also will try to help you to *evaluate your evaluations*. This is known as meta-evaluation and will be the subject of the last section of the handbook.

I. PLANNING COURSE EVALUATIONS

In our attempt to help you plan course evaluations, we have drawn heavily from two sets of evaluation guidelines. You may wish to become more familiar with each set. At the minimum, you should at least know that these exist. In the Introduction, we referred to the set of guidelines developed by the Joint Committee on Standards for Educational Evaluation. This Joint Committee, chaired by Daniel L. Stufflebeam, was composed of members from twelve national educational organizations. Their thirty standards (1981) are organized into four categories:

1. The Utility Standards are intended to ensure that an evaluation will serve the practical information needs of given audiences.
2. The Feasibility Standards are intended to ensure that an evaluation will be realistic, prudent, diplomatic, and frugal.
3. The Propriety Standards are intended to ensure that an evaluation will be conducted legally, ethically, and with due regard for the welfare of those involved in the evaluation, as well as those affected by its results.

4. The Accuracy Standards are intended to ensure that an evaluation will reveal and convey technically adequate information about the features of the object being studied that determine its worth or merit.

A second set of standards was developed by the Evaluation Research Society (ERS), an organization established in 1976 and chaired by Scarvia B. Anderson until 1980. Their fifty-five standards (1982) are organized into six sections:

1. Formulation and Negotiation
2. Structure and Design
3. Data Collection and Preparation
4. Data Analysis and Interpretation
5. Communication and Disclosure
6. Utilization.

In a comparison of these two sets of standards, Stufflebeam (1982) found that there was great overlap. According to our own synthesis of the Joint Committee and ERS standards plus based on our own professional experiences, basically there are seven steps necessary to planning a course evaluation which we will review one at a time:

A. Identify users and needs.
B. Ask specific questions.
C. Indicate how questions will be answered.
D. Select and/or design instruments.
E. Consider ethical issues.
F. Collect and analyze data.
G. Report results.

A. Identify users and needs.

Numerous individuals and groups potentially can benefit from a course evaluation. It is very helpful to identify all of these individuals and groups in advance for two reasons. First, it would be inefficient for *each* of these parties to conduct their own course evaluations. Second, it would be disappointing for *any* of these parties to not have their needs met by the course evaluation.

It strikes us that the likely users of a course evaluation include the course's director and teachers, the department head, the school curriculum committee, and the promotions committee. Whoever is spearheading the course evaluation plan (often the course director) should consult with representatives of the other parties to elicit their needs.

According to the classic model of program evaluation developed by Scriven (1967), there are two roles for evaluation, formative and summative. Formative evaluation is conducted as the process of teaching and learning occurs, and its purpose is to improve. Summative evaluation is conducted at the end of the teaching-learning process, and its purpose is to rate a course.

Improvement of medical school courses (formative evaluation) is a

never-ending process, and any one or all users of a course evaluation might be interested in formative evaluation. However, it would not be practical to conduct an extensive formative evaluation of each course in a medical school every year. In this regard, the Joint Committee warns that, "the evaluation should produce information of sufficient value to justify the resources expended (p. 60)." Also, the ERS recommends, "An estimate of the cost of the proposed evaluation and, where appropriate, of alternatives should be provided (p. 12)..."

In our view, new courses and revised courses deserve extensive formative evaluation. Also, courses known by reputation to be in trouble should be formatively evaluated. Otherwise, formative evaluation of courses can be minimal and, perhaps, periodic.

The rating of medical school courses also is never-ending because of the need for various users to make decisions. Summative evaluation can provide information which can be used in decision-making. For example, it can help...

- the course director make teaching assignments;
- the department head choose the course director;
- the curriculum committee decide whether to add a new course to the curriculum; and
- the promotion committee identify exemplary teachers.

In thinking about formative versus summative evaluation, evaluation users should keep in mind that the "conventional wisdom" is that the greatest benefits derive from the chance to make improvements and that summative evaluation often is overdone. *In planning course evaluation, it is essential that each user identify its information needs for making improvements (formative evaluation) and for making decisions (summative evaluation).*

B. Ask specific questions.

We find it helpful to frame each user's needs into specific questions which the evaluation will answer. When considering what questions to ask about a course, users might find it helpful to think about the "Three Ps": perceptions, process, and products. Each, alone, constitutes a limited view. It is helpful to know how people *perceive* a course, but what is liked may be neither meritorious nor worthwhile. It is helpful to assess the *process* of teaching, but we do not always know what promotes learning. It is helpful to know the *product* of a course in terms of learning, but assignment of causality can be confusing. We agree with McGuire's recommendations to suppress the fallacies of each approach as much as possible, use all three approaches, and link the results of each with the others (1973). This is known as triangulation (Guba and Lincoln, 1981, p. 106).

The three major groups whose perceptions should be considered are students, colleagues of faculty doing the teaching, and faculty doing the

teaching. Just as triangulation of perceptions, processes, and products may be advisable due to the shortcomings of each approach, triangulation of the perceptions of these three groups may be advisable due to the limitations of each group's view.

Some faculty believe that student perceptions are so limited that these should not be used in the evaluation process. Frankly, this is an "old saw." The fact is that student evaluation of instruction in higher education has been studied extensively. By and large, when surveys of students' perceptions are designed and used properly, students provide useful data. If you are skeptical, we refer you to an extensive literature review which justifies student evaluation *when used properly*. According to Overall and Marsh . . .

> A review of recent research concerned with student ratings of teaching indicates that such evaluations are not significantly influenced by background variables, and are valid, reliable, stable, generalizable, and useful. This research further indicates that properly constructed evaluation questionnaires assess multiple dimensions of the instructional process. Thus, when obtained from properly designed questionnaires, data from students' evaluations is particularly appropriate and useful in the instructional evaluation process (1982, p. 12).

Of course, we want to stress that the value of student ratings lies in their contribution to a broad data collection. We agree with Benton, who in his literature review of student ratings, concluded:

> It seems quite clear that students ratings of instruction provide good evidence of the quality of teaching. However, they provide evidence only; they should not be considered to be more than evidence. They should never be considered alone as positive proof. It is quite clear that there is something more to teaching than can ever be totally accounted for by those who are taught (1982, p. 43).

With regard to students' perceptions, simply knowing how they experienced a course might be valuable to evaluation users. Examples of questions which could be answered by students, identified from factor analysis studies of student rating forms, include (Centra, 1971, p. 23) . . .

1. Organizational Structure
   - Was material presented in an orderly fashion?
   - Was the instructor prepared for each class?
   - Was class-time well spent?
   - Was course well organized?
   - Did the instructor make clear what students were expected to learn?
   - Was there agreement between announced objectives and what was taught?
2. Teacher-Student Interaction or Rapport
   - Was the teacher available for consultation with students?
   - Did the teacher seem to know when students did not understand the material?
   - Did the teacher actively help students who had difficulty?

- Did students feel free to ask questions or express opinions?
- Did the teacher seem concerned with whether students learn the material?

3. Teaching Skill, Communication, or Lecturing Ability
   - Did the teacher use examples or illustrations to clarify the material?
   - Did the teacher present material clearly?
   - Did the teacher summarize or emphasize major points?
4. Workload and Course Difficulty
   - Was the workload heavy in relation to other courses?
   - Did the teacher try to cover too much material?
   - How difficult were the reading assignments?
   - Did the course challenge students intellectually?
   - Did students put a great deal of effort into the course?
5. Grading and Examinations
   - Did the teacher tell students how they would be evaluated?
   - Did the examinations reflect the important aspects of the course?
   - Did the teacher assign grades fairly and impartially?
6. Impact on Students
   - Did students believe they learned a great deal?
   - Did the course fulfill students' goals?
   - Did the course stimulate students to learn more in the same area?
7. Overall Ratings
   - How did the students rate the teacher's effectiveness?
   - How did students rate the worth of the course?
   - What was the general quality of classes?

Surely, in a given course evaluation, when users are asked what they want to know about student perceptions, not all the above questions will be seen as important. However, in some cases, more questions might be asked than would be feasible to answer. Therefore, evaluation planners should prioritize these questions.

With regard to colleague perceptions, it strikes us that peers provide a perspective that is particularly important, especially regarding issues which cannot be judged well by students alone or by students at all. Examples of questions which can be answered by colleagues include (Diamond, Sharp, and Ory, 1978)...

1. Content Suitability
   - Is the content worth knowing?
   - Does the content represent current thinking in the discipline?
   - Is the level of difficulty appropriate?
2. Organization of the Content
   - Is classtime used well?
   - Are the topics logically sequenced?
3. Clarity of Presentation
   - Does the teacher adequately explain new terms, concepts, and principles?
   - Does the teacher use relevant examples?

4. Questioning skills
   - Does the teacher ask a variety of types of questions?
   - Does the teacher encourage students to ask and answer questions?
   - Does the teacher respond adequately to questions?
5. Teacher's Style
   - Does the teacher stimulate students to think?
   - Is the teacher a good role model?
6. Teacher-Student Interactions
   - Does the teacher encourage student contributions?
   - Is there mutual respect?

In addition to student and colleague perceptions, self-assessment also can be a helpful data source. In particular, we recommend that course teachers pre-establish cues of successful instructional events. By identifying how students would respond in successful lectures, labs, and group discussions, the teacher can look for these cues and compare actual to desired student behaviors. Example of "success criteria" or cues include. . .

1. Lectures
   - Were students attentive?
   - Did they respond to questions?
   - Did they ask questions?
2. Group Discussions
   - Did student prepare adequately?
   - Were students actively engaged in discussion?
   - Did group members demonstrate respect for each other's viewpoints?
3. Laboratory
   - Did students follow instructions?
   - Did students complete activities?
   - Did students show enthusiasm for the tasks?
   - Did students seek appropriate assistance?

As we have pointed out, answering evaluation questions with perceptions alone is not advisable. A second major approach is to measure the process of teaching. In other words, we recommend that medical school teachers use educational specialists to analyze the whole instructional process. Examples of evaluation questions which can be answered by educational specialists include (McLagan, 1978). . .

1. What is done to help motivate students?
2. What is done to help students handle course information and experiences?
3. What is done to help students achieve cognitive, affective (attitudinal), and psychomotor objectives?
4. What is done to help students apply what was learned to practical task?

Over the past ten years, many medical schools have employed educational specialists. Sometimes, they are located in a single academic department, in Family Medicine, for example. Often, they are located in a

schoolwide office of educational research and development. In medical schools which support educational specialists, we recommend their use as resource people both in planning course evaluations and in providing their own measures of the teaching process. In medical schools which do not support educational specialists, we recommend that they seek out and support consultation from the school of education which may exist on the same university campus.

The third major approach to answering evaluation questions is to study the product, *i.e.*, learning outcomes. Using learning outcomes to evaluate a course and its teachers is controversial. Some educators do not favor using student achievement as a means to measure teacher effectiveness because of differences in the difficulty of courses and the potential abuse by instructors who "teach to the test" (McCarter, 1974, p. 32). Proponents argue that learning outcomes are the "ultimate criteria" of teaching effectiveness (Meeth, 1976, p. 4), and that when only the process of teaching is measured, only half the evaluation process is really accomplished (Mark, 1977, p. 104). Because we believe that whether or not students learned is relevant to any course evaluation, we recommend that products be used in conjunction with perceptions and processes. *Triangulation of major approaches to answer evaluation questions asked by evaluation users is a cornerstone to planning and implementing course evaluations which will serve the interests of a medical school.*

C. Indicate how questions will be answered.

Once the questions to be answered by the evaluation have been identified, it is necessary to indicate *how* the questions will be answered. Questions concerning perceptions of students often can be answered with questionnaires. Seyer and Smith (1976) point out that student questionnaires can gather information regarding opinions, interests, and attitudes:

Opinions. You may want to know students opinions about the course.
Interest. You may want to know topics are of most interest to students and what learning activities they prefer.
Attitude. You may want to know what attitudes you are helping students develop as you teach.

In reviewing this list, Knapper in a letter of correspondence to the authors, added the category of *Experience* to the list, i.e., you might want to know the students' experiences with the course. Often, student questionnaires are used at the end of a course. This approach can be helpful in both formative and summative evaluation. However, we can think of at least two instances when it also would be helpful to survey student perceptions *during* the course. First, when evaluating a new course or a new approach to an old course, formative evaluation might be faciliated by collecting student feedback on a on-going basis. Second, when evaluating a lecture-style course with many teachers, summative evaluation might be

facilitated by rating each lecturer when the student response is fresh. At the end of the course, it can be difficult for students to remember teachers who lectured early in the course. Even when *on-going* evaluation is not desireable or practical, sometimes a *mid-course* evaluation can be helpful, especially when specific improvements which can be incorporated immediately are solicited. Also, Knapper appropriately raises the issue of surveying student opinions, interests, attitudes, and experiences *at the beginning* as well as during and at the end of the course.

Keep in mind that questionnaires are not the only means to learning about student perceptions. Meeting with a student committee and discussing their views can be helpful. We strongly recommend this group discussion approach when the evaluation questions concern policy questions, e.g., should a course have more or less hours? When asked on a questionnaire, the student response becomes a mere poll. Knowing the *number of students* who want more, less, or the same hours will not likely help decisions makers. Knowing the *student reasoning* for wanting more, less, or the same hours can help decision-makers. In these cases, it is the thought that counts!

Questions concerning colleague perceptions of classroom teaching are not best answered by questionnaires. Our experience is that faculty prefer to discuss their views in a meeting. Some faculty feel uncomfortable with observing other faculty or with being observed by other faculty; so the use of this approach should be discussed carefully by the evaluation users during the planning stages of evaluation.

Some faculty feel that observation of one or two of their lectures by a colleague provides an unrepresentative view of their teaching. Instead, they prefer colleagues to look at the overall approach taken in a course. Thus, a new approach, which is more acceptable to some faculty, is the use of peer review to evaluate teaching dossiers. For faculty interested in the use of teaching dossiers to improve the quality of teaching (formative evaluation) or to include teaching effectiveness among the criteria for making personnel decisions (summative evaluation), we strongly recommend the *Guide to the Teaching Dossier: Its Preparation and Use*, prepared by the Teaching Effectiveness Committee of the Canadian Association of University Teachers. According to this guide (Shore et al., 1980)...

> The dossier approach to recording teaching accomplishments is intended partly to forestall compulsory teaching evaluation by questionnaire alone. Teaching is an often creative activity whose success is shown by a variety of data from different sources (p. 3).

The teaching dossier, as developed by the Canadian Association of University Teachers, was proposed as a general strategy of evaluation, placing responsibility on the individual teacher to record teaching activities and successes. Although reviewed by colleagues, the contents of the dossier come from a variety of sources. Their list of forty-five items which

could be included in a teaching dossier are shown in Appendix A of this handbook. Also, French-Lazoviks's form which can be used by colleagues to rate the materials in the teaching dossier is shown in Appendix B.

When using educational specialists to evaluate the process, you should expect them to provide a plan for their portion of the evaluation. In a sense, the educational specialist is a "consultant" and you are the "client." Each role carries with it mutual responsibilities. Both sets of evaluation guidelines provide standards which are relevant to the consultant-client relationship:

> The persons conducting the evaluation should be both trustworthy and competent to perform the evaluation, so that their findings achieve maximum credibility and acceptance (Joint Committee, 1981, p. 24).
>
> Evaluators should not accept obligations that exceed their professional qualifications or the resources available to them (ERS, 1982, p. 13).

When answering questions which concern products, i.e., learning outcomes, the evaluation plan should identify the use of internal and/or external examinations. Internal exams are those developed by the medical school to test the students. External exams are those standardized exams developed outside the medical school, for example the National Boards of Medical Examiners, licensing exams, and board certification exams.

Examinations, internal or external, can be helpful in rating a course; hence, these can be used mostly to answer summative evaluation questions. To answer formative evaluation questions, feedback from students and teachers regarding student performance later in the curriculum can be helpful. Asking alumni for feedback regarding their medical school course provides an additional help in improving medical school courses.

When identifying questions and indicating how to answer these questions, the evaluation planners need to be realistic regarding the burdens placed on students and faculty. In this regard, both sets of evaluation guidelines provide identical advice: disruption to course activities must be kept to a minimum (Joint Committee, p. 52; ERS, p. 14).

Thus, a major challenge of the evaluation plan is to provide satisfactory answers to evaluation questions without demanding time and effort disproportionate to the task.

D. Select and/or design instruments.

Selecting instruments which already exist and designing your own instruments constitute a critical step in the planning process because the instruments limit the *formal* data collection. Informally, students and colleagues might volunteer information about a course, and informal data can be quite useful. However, formally, you will only know what you ask. Since, as we already have pointed out, evaluation has to be practical, what you ask should be carefully chosen.

In this handbook, we are focusing on student questionnaires.

Although we recommend that colleague perceptions be measured through discussion of classroom visits and/or by their review of teaching dossiers, should you choose to design colleague questionnaires, the principles of designing student questionnaires will be applicable. When you decide to evaluate the process, we recommend that you ask your educational specialists for help with selecting and designing instruments. The need for their help is explained by French-Lazovik (1982):

> Evaluating a complex human performance like teaching requires forms of psychological measurement that pose difficult technical problems. Faculty members, for the most part, have enough awareness of these problems to ask, "Are the data collected reliable and valid?" But the conditions likely to produce psychometrically sound results and the methods of analysis that reveal the degree of reliability and validity present in data are not general knowledge. They are the special domain of the expert in psychological measurement (p. 45).

In addition to *not* addressing measurement of process in this handbook, we also will not discuss measurement of products, *i.e.*, learning outcomes. The issues of test construction are numerous and complex. In addition to asking for the assistance of educational specialists, we recommend Clark's chapter on "Measures of Student Learning" in Centra (1981).

With regard to student questionnaires, you should be aware that a number of standardized instruments are available, *e.g.*, the Student Instructional Report (SIR) has been developed by the Educational Testing Service and the Instructional Development and Effectiveness Assessment System (IDEA) has been developed by Kansas State University. A standardized instrument developed outside your medical school can be useful when you want to conduct summative evaluation *and* want to avoid the task of developing your own. Personally, we recommend that a medical school develop its own questionnaire to conduct summative evaluation of its courses. We also suggest that the format permit additional questions formulated by the course teachers. By including their questions, the questionnaire can serve both formative and summative purposes. In addition, when the need for formative evaluation is great, *e.g.*, when a course is new or when an old course has been revised, a questionnaire tailored for the course should be considered.

The most common format for student questionnaires is the *rating form*. Basically, a rating form is an instrument that requires the observer to assign the rated objects to categories on a numerical continuum (Kerlinger, 1964). As Anderson *et al.* (1973) point out, ratings essentially are subjective assessments made on a established scale.

Thus, there are two key elements of the rating form: the objects to be observed and the rating scale to be used. As we pointed out in Step B ("asking important questions"), appropriate objectives for student surveys include:

1. organization structure
2. teacher-student interaction
3. teaching skills
4. course workload
5. grading and examinations
6. impact on students
7. overall ratings

In other words, student questionnaires should focus on how students *experienced* the course, not on the content of the course or the expertise of the faculty.

With regard to format, we like the advice of Seyer and Smith (1976):

1. Put only one thought in each short, concise statement.
2. Use conversational words and phrases.
3. Avoid cliches.

With regard to rating scales, a general rating scale, *e.g.*, excellent to poor can be useful. Also, Likert scales, *i.e.*, strongly agree to disagree, are popular. In either case, you must decide whether to use an odd or even scale. A odd scale provides a neutral rating; whereas, an even scale forces a positive or negative rating.

Regardless of your choice, we agree with Irby and Dohner's recommendation for no fewer than three positions on a scale and no more than seven (1976).

For your perusal, two rating forms are included in the Appendix. Appendix C, which uses a Likert scale, is a form designed to provide formative evaluation of a single lecture. Appendix D, which uses a general rating scale, is a form designed to provide summative evaluation of a teacher. In addition to rating scales, some open-ended questions can be useful, *e.g.*, "How do you think the course can be improved?" Open-ended questions are particularly helpful in formative evaluation. The responses to these questions are vulnerable to subjective interpretation and, in general, should not be used in summative evaluation of teachers. In particular, you should guard against making judgments about a teacher based on a few highly negative or positive remarks.

What we want you to remember regarding student questionnaires is to use items which will tell you how they experienced the course, and, if you used a rating form, a scale which will summarize their experiences.

E. Consider ethical issues.

Course evaluation can be threatening for at least two reasons. First, it asks for *judgments* to be made regarding faculty performance. Second, it asks for *decisions* to be made regarding teaching assignments and appointments. Society-at-large is concerned with how data are collected and used, so it should not be a surprise that course evaluation will raise many concerns in the medical school.

Both sets of evaluation guidelines emphasize the importance of ethical issues. In fact, "propriety standards" comprise one of the Joint Com-

mittee's four major categories of standards (1981), and the ERS identified ethical issues in each of its six major sections (1982).

According to the ERS, restrictions, if any, on access to data and results from evaluation should be clearly established and agreed to at the outset. For example, your policy could be that only the department head and the faculty who are teaching a course should have access to data and results from evaluation conducted for formative purposes, *i.e.*, to improve teaching and the course-as-a-whole. On the other hand, the curriculum committee and promotions committee may need access to data and results from evaluation conducted for summative purposes, *i.e.*, to make curriculum decisions about the course and to make personnel decisions about the faculty.

Also, according to the ERS, all data collection should be conducted so that the rights of individuals are respected and protected. This standard raises the question of whether students ratings should be anonymous. By keeping student ratings anonymous, more weight is given to protecting students' rights than to teachers' rights. Although it can be argued that students will give more thoughtful ratings if they are required to sign their names, in our view, anonymous ratings are preferable.

Finally, according to the ERS, evaluation findings should be presented fairly. To meet this standard, we recommend that the author of a course evaluation report share a draft with course teachers and elicit their interpretations of the data. When there is disagreement over interpretation, the final report should present the different views and identify their sources.

One ethical issue always to be considered is whether staff involved in the formative segments of the course evaluation should be involved in its summative segments. The question is whether a person who is dedicated to helping faculty improve can objectively evaluate faculty for the purpose of making decisions. When, because of limited resources, the formative and summative evaluators are the same, awareness of the potential bias should at least provide a note of caution.

We want you to keep in mind that, in planning an evaluation, the welfare of both those being evaluated and those providing evaluative data must be considered.

F. Collect and analyze data.

As consumers of the educational product, students are able to provide valid and reliable information concerning many of the elements of good instruction. Students are not content specialists and should not be used as such, but student views concerning organization, integration and usefulness of course materials has consistently been shown to be reliable and valid. Information is defined as reliable if the same results ensue over repeated measures. A measure is valid if it is able to measure the area of interest for which it was designed. It has been shown repeatedly that

students are reliable and valid sources of information concerning course organization, clarity and integration.

Students can be used on course education committees that meet periodically to evaluate course progress. Many faculty feel that the feedback provided by students on the course education committee is invaluable. Most course education committees meet several times a week to review the progress made toward accomplishment of course goals. Students are able to indicate if the instructors are moving at an appropriate pace, if examinations were fair and accurate assessments of materials presented, if handouts and reading assignments were overly easy or difficult, and provide information regarding specific lecturer technique that might require improvement, e.g., "The lecturer moved too rapidly through the material," or, "The lecturer was disorganized and ill prepared." Such information can assist course leaders in better adapting to the audience.

Selection of students for course education committees can best be accomplished by a request for volunteers. Students who have a genuine interest in the course will most probably step forward and the interest and enthusiasm of such students can be a valuable resource to the instructors.

It should be kept in mind by instructors that such student input as provided on the course education committee is potentially biased. If feedback from students on the committee is alarming, the education committee may be advised to undertake a more representative sampling of the entire class to confirm if the information received is so viewed by the students at large.

It has also proven useful for the course director to take notes during the course education committee meetings. It not only serves to document that such meetings were held, but may provide useful information for the future. Faculty, being of the normal, fallible, human variety, can sometimes forget past mistakes and the committee minutes can help avoid pitfalls in subsequent courses.

It is very important that the course director communicate the importance of thoughtful and objective evaluation of the faculty and course to the students. The course director will receive a much more favorable response from the students when they view their feedback as important and seriously considered. Secondly, the evaluation should be administered at a time during class when students have adequate time to make thoughtful responses. The evaluation should not be administered either immediately before or after an examination and it is to the advantage of the faculty if the number of questions asked on the evaluation is kept to a minimum, e.g., limit the evaluation to a single section of the course and ask questions that will serve to give the instructor or instructors information that is short, concise and to the point. Most medical schools employ education specialists who possess expertise in writing of such evaluations and their assistance should be sought. It is also advisable that students are

permitted an area for commenting upon the section as well as for making ratings. Typically asked questions in the comments section are "What did you like best about this section?" and "How could this section be improved?" The same questions could be asked regarding the instructors. A sample formative evaluation of a course section is provided in Appendix E.

With regard to sampling, it is best to include all members of a student class. A response rate of 75% is considered reliable. If it is not practical to survey an entire class, a random sampling of twenty percent to tthirty percent is reasonable when a number of formative evaluations are conducted in a course. In such instances, a table of random numbers can be used to select a given portion of students for the evaluation. Your sampling technique should be carefully explained to the students and their cooperation sought. If a random sample is to be obtained, each student should be identified in order to assure that all evaluations are completed in the event that a follow-up is required. Again, students should be assured that their names or identification numbers are used only to assure a complete return of evaluations and that the names or numbers will be destroyed so that faculty will not be able to identify a student name with an evaluation. Many students are reluctant to provide candid evaluations of instruction if they believe the instructor or instructors will be able to identify them, and every effort should be attempted to assure the students that the evaluations will be treated with strictest confidentiality. It is oftentimes advantageous to recruit the assistance of a member from the Office of Medical Education in the medical school to administer the evaluations to further remove faculty from access to identifiable evaluations. The Office of Medical Education is then able to summarize the information, and serve as an intermediary between faculty and students on this very sensitive issue. While some faculty view the precautions with skepticism, the goal is to obtain useful information from students regarding the course and if such precautionary measures are necessary, it seems a small price to pay.

In summary, formative evaluations should be short and to the point. Lengthy, complex evaluations are to be avoided. Students should be permitted time during class to complete the evaluation and solicit cooperation. It is inadvisable to administer the evaluation before or after an examination as undue bias may result. Educational specialists should be sought to provide assistance in construction of the evaluation form.

Much of what was written on formative evaluation applies to summative evaluation. The evaluation form should be short, concise, and easily answered. Educational specialists should be consulted in design of the instrument and student input is also advisable. The evaluation should not be administered at a time when students would be rushed, such as before or after an examination and the course director should stress the importance of thoughtful response to the items on the form.

The evaluation instrument can be designed for computer scoring and this saves much time and effort. One popular method for accomplishing computer scored evaluations is for students to mark the responses on an optical scanning sheet that can be scanned by a special optical scanning instrument common to most computer centers. Once the sheets are scanned and read onto tape or disk, the information can be "massaged" by various statistical software to generate frequency distributions, mean ratings, and response frequencies. When the evaluation, either summative or formative, has been administered, the frequency and percent of responses should be generated for each item on the evaluation and a mean value calculated. This information should be displayed in the summary report alongside each evaluation question. These data serve to provide the instructor with general information concerning how well the course was received. An example is provided in Appendix E.

Faculty, of course, are the backbone of the course education committee. It is this committee that is responsible for the design and implementation of the course and faculty input is indispensable. Unfortunately, many such committees discontinue meeting after the course is underway and an opportunity for further discussion and evaluation is lost. In many instances, however, these committees continue to meet to review progress and assess the impact of the instruction and materials on the students. Some education committees go so far as to videotape specific lectures or demonstrations and critique them for improvement. These episodes, while possibly painful for the individual videotaped, are a source of incalculable influence for good. In these sessions, the lecturer or demonstrator receives firsthand feedback from his peers, and this feedback, can have a great influence on improvement of instructional skills.

Even if individual presentations are not videotaped, it is beneficial for the teaching faculty to meet periodically to review progress made toward achievement of the instructional goals. It is essential to the proper course function, especially if the course is long in duration.

In summary, peer review on the course education committee can be a very positive and rewarding experience to those involved. There aren't many who take joy in the review by peers, but if done considerately, this information can change the attitudes and skills of a very mediocre instructor into those of a highly competent, respected educator of medical students.

Soliciting the support of teaching faculty in review of course and instructors is oftimes difficult. Medical school faculty are burdened with multiple responsibilities, and asking that they take the time out of busy schedules to attend a series of lectures or demonstrations is sometimes asking more than is reasonable. However, if faculty are available and willing to take the time to undertake an evaluation of another peer, the information obtained is invaluable. No other source is available for review of the content matter of the lecture or demonstration and it is for this purpose that peer review is essential.

The educational specialist can serve as a valuable resource on course education committees. Such individuals posses skills that are very useful in development of course objectives, examination construction, syllabus preparation, and for providing advice to assist faculty in refinement of lecture and discussion skills. The educational specialist is also a useful source of assistance in construction of formative and summative evaluations and if it is possible to enlist the services of a technologist the faculty and students usually benefit greatly from his services.

Course examinations are rich in information. The following statistics should be obtained to assist the instructor in evaluation of the examinations impact: the mean, standard deviation, number of students examined, lowest and highest scores. Each of these scores should be expressed in raw score form and as a percent of total points possible. In addition, the instructor should be provided with a distribution of scores, an alphabetical roster of student scores expressed as raw scores, percent and rank and roster of scores ranked from high to low.

Each student should receive a report detailing his raw score and percent correct, class average and standard deviation, (expressed as raw scores and percents) and the student's rank in the class. Included in the student report should be a listing of each examination question, the correct answer and the student response if his response was incorrect. This form of feedback is very useful to the student and instructor alike, as it informs each as to where the student stands and identifies content areas in which the student is poorly prepared.

For each examination question the instructor should be provided with the frequency and percent of students who selected each alternative if the questions are true — false or multiple choice. An item-total (Point Biserial) correlation should also be generated for each question. This correlation gives the instructor general information regarding the ability of the item to discriminate, i.e., if students with high test scores choose the correct response and students with low test scores select an incorrect response, the item is said to discriminate. Point Biserial correlations in the .40 range or higher indicate that the item can discriminate. Conversely, and more importantly, if the value is zero or negative then the question may be poorly written, or the students simply were guessing. This correlation should be used more as in indice of poorly written or ambiguous questions and can give the instructor valuable data concerning the questions themselves.

One further statistic should be obtained on each subsection of the examination, coefficient Alpha. The coefficient reflects how well the test questions are interelated in assessing a particular domain of knowledge. If one is testing a specific domain of knowledge, e.g., fluid and electrolyte balance, then the questions should be homogenous and directed to that area of concern. The coefficient Alpha shows if the items are related and if there are a sufficient number to adquately access that particular area.

Alpha coefficients should be at least in the .80's and preferably higher.

The instructor should make every effort to collect reliable and valid information in order to analyze the impact of the teaching effort on student learning.

G. Report results.

The information obtained from a formative evaluation completed by students should be reserved for the course director and faculty directly involved with teaching. It is not usually appropriate for this information to be shared with administrators, curriculum committees or other administrative bodies peripherally involved with the course. The expressed purpose of formative evaluations is to assist faculty in skill improvement and course modification and should not normally be used in administrative decisions.

The data generated from a formative evaluation should be summarized and presented in a fashion that is easily read and understood. If numerical ratings are included as part of the evaluation, the frequencies, percents and item means should be displayed as seen in Appendix E.

The reporting of summative evaluation information should be designed to suit the needs of the audience for which the report was intended. For example, the department chairman would receive a report different from that prepared for a curriculum committee and teaching faculty should receive the most comprehensive summary of findings.

As with formative evaluations, summative findings should include item statistics such as rating frequencies, percents and means as seen in Appendix E. If written comments are included, they should be subsumed under topic headings. *The results of a summative or formative evaluation should be reported in a fashion that provides the reader with complete and accurate description of the information obtained.*

PART II. USER-CENTERED EVALUATION

A. Evaluation for Course Directors and Teachers

The director of a course and the faculty who teach in it are obvious users of course evaluation. Sometimes, a course director teaches alone or with only a few other faculty. This may be the case in an elective course for medical students or a seminar-course for a small group of residents. When only a single or a few faculty are involved in directing and teaching a course, their evaluation needs may be relatively simple. Their need to collect student feedback and to make improvements often can be satisfied by informal evaluation and limited formal evaluation.

Course evaluation is more formal when there is a course committee rather than a single course director and when there are a large number of teachers rather than one or a few. Albanese *et al.* (1979) report an increase in multiple instructor courses in both professional and graduate schools due to the increase in specialization in all fields.

The use of many instructors in one course raises the issue of whether students should be questioned immediately after a lecture or at the end of the course. To avoid the dependence of instructor ratings on general memory (pointed out by Thornedike and Hagen), Albanese *et al.* (1979) randomly divided 199 sophomore medical students into seven evaluation groups and asked each group to assess instructors for a two-week period in the Sophomore Introduction to Clinical Medicine course at Iowa. Because this course included 333 lectures by 175 different lecturers, the evaluation committee recognized that requiring each student to complete a separate questionnaire for each lecture or lecturer would be too great a burden. On the other hand, they anticipated that a single questionnaire at the end of the course would be too confusing for students. By dividing the students into evaluation groups, the course evaluation achieved an 86% student completion rate.

The issue of immediate versus end-of-course student questionnaires also was addressed by Irby *et al.* (1977). In the Sophomore Introduction to Patho-psychology course at Washington, 13 lecturers delivered 14 lectures (one teacher gave two lectures). The evaluation committee divided the lecture hall into four sections and distributed rating forms to one section for each lecture. The intent was to give out 30 rating forms daily, roughly one-fourth of the class. The mean completion rate was 80%. In addition, at the time of the final examination, each of the 14 lectures was rerated by one-fourth of the class. This was done by asking each student to think back and rerate four randomly chosen lectures using rating forms identical to those used during the daily ratings.

Because of the stability of immediate and postcourses ratings, Irby *et al.* (1977) recommend that decision of when to question students should be determined in part on the basis of faculty needs. If there is a need for constructive feedback and course improvement, immediate evaluation provides more specific comments. If there is a need to rank faculty and courses, postcourse evaluation appears to be reliable. Of course, reliability of student questionnaires will decrease as the number of different faculty increase.

Determining when to administer a questionnaire is one issue. Selecting and designing a questionnaire is another. Albanse *et al.* (1979) selected 16 items from the Student Perceptions of Teaching (SPOT) question pool developed by Whitney and Weerts (1974) and added four items developed by course personnel to meet special needs of the course. We suggest that course directors ask for samples of standardized student questionnaires from their Office of Medical Education and that they feel free to tailor questionnaires to meet their own needs, as was the case in Iowa. The criteria for question selection at Iowa, which we applaud, were:

1. a statement in terms of the individual student's perception as distinct from the student's judgment for the class;

2. reflection of a single descriptive characteristic or behavior to avoid multiple interpretations;
3. reference to a characteristic which the student could be expected to validly observe; and
4. applicability to the majority of the instructors in the course as opposed to a subset.

We strongly recommend that course directors use these criteria developed by Albanese *et al.* at Iowa to judge each item on a student questionnaire.

In the first chapter on planning course evaluations, we suggested that educational specialists in your medical school can be used to help answer questions about the process of teaching. When faculty in a course feel a need for individual treatment plans to improve their teaching skills, educational specialists can provide the desired prescription. Irby *et al.* (1976) described the development of a lecture observation tool in which 13 instructional variables and four personal attributes were included (see Appendix F).

To provide meaningful feedback to the instructor, the educational specialists from the University of Washington School of Medicine, Office of Research in Medical Education (ORME) recorded concise statements describing performance rather than a rating scale. For example, for the item, "stimulates learner questions, responses, and discussion," the observation, "you missed responding to student questions on four occasions," was deemed more helpful than a rating of "poor" or "needs improvement."

The procedure at Washington was for the ORME observer and instructor to meet for 30 minutes immediately after lecture. The ORME consultant sought to involve the lecturer in self-evaluation as well as to discuss the consultant's observations. We want to emphasize that, if the goal of evaluation is to help individual teacher's improve their teaching performance, *shared analysis* as soon after the event as possible is essential.

We think that the experience of Irby *et al.* (1976) in using an educational specialist to evaluate teaching for the sake of improvement is valuable and should guide your efforts:

- An evaluation program to help teachers in a course to improve should be tailored to fit the particular situation.
- Evaluation consultants should recognize that junior and senior faculty have different individual needs.
- Course instructors should be involved in both the planning and evaluation phases.

When planning a course evaluation, the course director and faculty should consider the ethics of protecting data sources. If a student questionnaire is going to be used, the students must be assured that all

responses are confidential. Some faculty complain that, when student questionnaires are anonymous, some students respond irresponsibly. Our experience is that, at most, a very few students abuse the anonymous source of questionnaires by making irresponsible statements. On the other hand, by protecting the source of questionnaires, we find that students feel they can be honest. Also, by assuring that all responses are anonymous, more students are willing to respond.

Other ways to increase the response rate include giving a list of course changes from the previous year which were made on the basis of student evaluation (Albanese *et al.*). An example of improvements made in a course based on student and peer evaluation is provided by Stillman *et al.* (1983), who documented improvement in individual instructors and the overall course in a two-year study of the Clinical Correlations with Pathology Course at Arizona.

Stillman *et al.* suggest that overall improvement in the course as perceived by students was explained by improvement in faculty performance due to feedback and participation by different lecturers. In multi-instructor courses (the Arizona course used 50 different lecturers), the course evaluation can help the course director separate the wheat from the chaff. Teachers with highly favorable evaluation results should be thanked profusely and encouraged to continue to participate in the course in the future. For teachers with good, but not great results, the course director should express gratitude and decide on a case-by-case basis, whether it is worthwhile to help the good teacher become better. For teachers with unfavorable evaluation results, the course director must decide whether to accept the teacher as is, try to change the teacher's performance, or simply exchange the teacher for someone else. We readily acknowledge that these decisions are political as well as pedagogical.

Our experience is that the better someone teaches, the more open the person is to making improvement. Resistance to change raises the ethical issue of whether faculty automatically or voluntarily are subjected to evaluation. One view, which we support, is that all faculty should be automatically included in end-of-course evaluations. However, faculty should be included in ongoing evaluation on a voluntary basis. Our thinking is that, if a faculty member does not want immediate feedback after a lecture, evaluation will not be a positive experience.

B. Evaluation for Department Heads

The department head also is an obvious user of a course evaluation. While it may be feasible to assign overall responsibility for a course to a course director, the department head ultimately is accountable for the quality of the instruction and the learning outcomes. The delegation of responsibility for a course by a department head to a course director raises the first evaluation need for a department head — the course

evaluation can be used to evaluate the performance of the course director.

The *formative* question which the course evaluation can answer includes "How can my course director do a better job?" Based on what has been learned in the course evaluation, the department head has an opportunity to give feedback to the course director which will improve that person's future performance as a course director.

The *summative* question which the course evaluation can answer includes, "How good a job did my course director do?" Based on what has been learned in the course evaluation, the department head can make a informed decision to re-assign responsibility to the same faculty member or to another one. Also, the department head can reward a course director with praise and recognition. In addition, this is an opportunity to document in the course director's own promotion file how well these responsibilities were carried out.

In order to answer these formative and summative questions, a department head has to insure that the course evaluation collect data from students and faculty regarding the organizational structure of the course and the grading and examination policies. Specific items to be included on rating forms are suggested on page 4 and 5.

When a department head checks out that these items are included on the rating forms, an ethical issue to be considered is to let the course director know that these results will be used by the department head to evaluate him/her. Also, the department head must consider discussing with the course director the weight given to informal feedback from course teachers and whether this feedback will be elicited by the department head.

Another source of data is the feedback to the department head provided by a student committee. Again, the ethical issue of clearing this with the course director and whether to include him/her in the meeting with students needs to be considered.

Aside from evaluating the performance of the course director, the evaluation needs of the department head are identical to those of the course director. Thus, we recommend that department heads review the course director's plan.

C. Evaluation for the Curriculum Committee

With respect to evaluation information, the Curriculum Committee requires such data for a variety of reasons. The Committee may wish to gather formative information to determine if a course is progressing satisfactorily. This is frequently the case with new courses or courses in which problems have been identified. The Curriculum Committee may also be interested in obtaining periodic summative data to evaluate the overall medical school curriculum or may choose to assess specific blocks of the curriculum to determine, for example, if students are being ade-

quately prepared for clinical experiences or postgraduate training. The Curriculum Committee may also wish to evaluate the efficacy of specific teaching formats such as organ system teaching as opposed to discipline oriented instruction.

In gathering the necessary information, the Curriculum Committee typically proposes specific questions it wishes answered, such as, "Are students adequately prepared in the basic sciences?", "Are our students adequately prepared for residency training?", or "Is organ system instruction more effective than discipline oriented instruction?" Whatever the questions may be, they are usually discussed at length before information is sought.

Once the questions have been formulated, the Committee is then free to determine the best sources to be considered in obtaining the answers. Not infrequently expert opinion is sought both from within and without the medical school. If classes are progressing unsatisfactorily, the teaching faculty might be interviewed to better understand the problems. In addition to personal interview and discussion the Committee may choose to review available formative and summative evaluation information to gain further understanding. In questions pertaining to the adequacy of preclinical or clinical instruction, the National Board Part I or Part II scores may be examined in order to compare the cognitive mastery of the school's medical students with students nationally.

In many instances medical school graduates are surveyed to obtain the residents' views concerning adequacy of preparation for the responsibilties incumbent upon the resident. Administrators and faculty of the resident's training program might also be contacted to assess their views regarding how well the particular school's graduates are prepared. Within a medical school, clinical faculty might be interviewed or surveyed by the Curriculum Committee to assess if students are adequately prepared in the basic sciences for introductory clinical experiences, and individual courses or departments are sometimes reviewed to assure that instruction is adequate.

A number of sources can be drawn upon to gather the essential information, but it should be kept in mind that information concerning individual students, courses or faculty should be treated with great respect and confidentiality. Sometimes the information reviewed by the Curriculum Committee is highly sensitive and every effort should be taken to ensure it will be treated with appropriate respect and consideration.

D. Evaluations for the Promotions Committee

Committees which deal with promotion, retention or tenure of faculty require reliable and valid information concerning a faculty member's contribution to the fund of knowledge, and his effectiveness as either a clinician or teacher or both. Research publications are typically used to

assess one's contribution to the field of science; patient load and peer review provide information concerning clinical skills, and student and peer evaluation are submitted as evidence of teaching effectiveness. The Promotions Committee should discount information that might be unreliable or invalid. For example, if the person under consideration has taught for several years, all information should be reviewed and if sampling of student input was insufficient or biased it should be discounted. The Promotions Committee should be able to assess the type of teaching undertaken, i.e., lecture, lab instruction, small group discussion, seminar leader or clinical instruction on rounds and conferences, and it should be satisfied that the information reviewed accurately reflects the faculty member's skills.

The Promotions Committee must be able to determine if a faculty member should be retained, promoted or be granted tenure and the import of these decisions demands trustworthy information. The Committee should rely on expert opinion from peers and educational specialists when forming judgments concerning the value of student evaluation of instruction. It should make every responsible effort to weigh the information from an ethical and unbiased perspective.

PART III. META EVALUATION

We strongly recommend that you *evaluate your evaluations*. Adopting criteria developed by Stufflebeam (1978), we suggest that you assess four areas of your course evaluations.

A. Usefulness

A *sine qua non* of good course evaluation is its usefulness. If some person or group does not find the course evaluation helpful in some way, we would question *why* it was done in the first place?

To assess the usefulness of a course evaluation, ask yourself...

Were meaningful questions asked and answered?
Was the reporting timely?
Was the report credible?

B. Practicality

A course evaluation should not take time and effort greater than the task of teaching. If the demands are too great, faculty and students will stop cooperating and the evaluation will not be implemented as planned. Also unrealistic expectations on people will deter them from cooperating in the future. So, be realistic and do what is feasible.

To assess the practicality of a course evaluation, ask yourself...

Was student time respected?
Was faculty time respected?
Were support services available?
Were the costs reasonable?

C. Technical Adequacy

Educational specialists in your institution should be used to insure that course evaluations are conducted at a level expected given the state-of-the-art of evaluation methodology. We suggest being wary that these specialists not be allowed to go too far in upgrading the standard of evaluation. Your purpose is to evaluate an educational program, not conduct educational research. Also, keep in mind, "an evaluation not worth doing is not worth doing well." (Patton 1980).

To assess the technical adequacy of a course evaluation, ask yourself...

Was the overall plan comprehensive?
Were the results reliable and valid?
Were multiple data sources used?

D. Ethics

Patton (1980) argues that, "Issues of evaluation methodology are issues of strategy, not of morals. Purity of method is no virtue" (p. 17). We tend to agree. In terms of technical adequacy, there is no moral obligation to a golden standard of course evaluation. However, there are moral obligations to treat all parties fairly — those providing data as well as those being evaluated. Because any evaluation tends to be biased in some way, meta evaluation plays a key role in controlling this bias.

To assess the ethics of a course evaluation, ask yourself...

Were student data anonymously collected?
Was peer review confidential when appropriate?
Was the objectivity of evaluators enhanced?
Were results appropriately reported?

Good luck with your course evaluations. We hope this handbook will help you *plan, implement,* and *evaluate* your course evaluations.

REFERENCES

1. Albanese, Judy *et al.* "Student Perceptions of Instruction: Assessment in a Multiple Instructor Course Which Maximizes Student Response." *Evaluation and the Health Professions* 2 (231-239) 1979.
2. Anderson, S.B. *et al. Encyclopedia of Educational Evaluation.* San Francisco: Jossey-Bass Publishers, 1973.
3. Benton, S.E. *Rating College Teaching: Criterion Validity Studies of Student Evaluation-of-Instruction Instruments.* AAHE-ERIC/ Higher Education Research Report No. 1, 1982.
4. Centra, J. *Determining Faculty Effectiveness.* San Francisco: Jossey-Bass Publishers, 1981.
5. Diamond, H., Sharp, G., and Ory, J.C. *Improving Your Lecturing.* Urbana: University of Illinois Office of Instructional Resources, 1978.
6. Evaluation Research Society, "Evaluation Research Society Standards for Program Evaluation," in *Standards for Evaluation Practice,* Peter H. Rossi, Editor. *New Directions for Program Evaluation* 15: 1982.
7. French-Lazovik, G. "The Role of an Evaluation Specialist." in *Practices that Improve Teaching Evaluation,* Grace French-Lazovik, Editor. *New Directions for Teaching and Learning* 11: 1982.
8. Guba, E. G. and Lincoln Y.S. *Effective Evaluation.* San Francisco: Jossey-Bass Publishers, 1981.
9. House, E. R. *Evaluating with Validity.* Beverly Hills: Sage Publications, 1980.
10. Irby, D. *et al.* "A Model for the Improvement of Medical Faculty Lecturing." *Journal of Medical Education* 51 (403–409): 1976.
11. Irby, D. *et al.* "The Use of Student Ratings in Multiinstructor Courses." *Journal of Medical Education* 52 (668–673) 1977.
12. Irby, D. H. and Dohner, C. W. "Student Clinical Performance," in *Teaching in the Health Professions,* C. W. Ford and K. Morgan, Editors. St. Louis: C. V. Mosby, 1976.
13. Joint Committee on Standards for Educational Evaluation. *Standards for Evaluations of Educational Programs, Projects, and Materials.* New York: McGraw Hill Book Company, 1981.
14. Kerlinger, F. N. *Foundations of Behavioral Research.* New York: Holt, Rinehart, and Winston, Inc., 1964.
15. Mark, S. F. "Faculty Evaluation Systems: A Research Study of Selected Community Colleges in New York State." New York: Faculty Council of Community Colleges, State University of New York, 1977.
16. McCarter, W. R. "Making the Most of Subjectivity in Faculty Evaluation." *American Vocational Journal* 49 (32–33) 1974.

17. McGuire, C. "The Evaluation of Teachers and Teaching Effectiveness." World Health Organization Publication WHO/Educ/73. 165. Chicago: University of Illinois Center for Educational Development, 1973.

18. McLagan, P. C. *Helping Others Learn: Designing Programs for Adults*. Reading, Massachusetts: Addison-Wesley Publishing Company, 1978.

19. Meeth, L. R. "The Stateless Art of Teaching Evaluation." *Change Report on Teaching* 2(3–5) 1976.

20. Overall, J. V. and Marsh, H. W. "Students' Evaluations of Teaching: An Update." *AAHE Bulletin* 35, 1982.

21. Popham. W. J. *Educational Evaluation*. Englewood Cliffs, New Jersey: Prentice-Hall, Inc., 1975.

22. Rippey, Robert M. *The Evaluation of Teaching in Medical Schools*. New York: Springer Publishing Company, 1981.

23. Scriven, M. "The Methodology of Evaluation," in *AERA Monograph Series on Curriculum Evaluation*, No. 1, Robert E. Stake, Editor. Chicago: Rand McNally, 1967.

24. Seyer, P. C. and Smith, C. R. *Developing Opinion, Interest, and Attitude Questionnaires*. San Jose: San Jose University Faculty and Instructional Development Office, 1979.

25. Shore, B. M., *et al. Guide to the Teaching Dossier: Its Preparation and Use*. Ottowa: Canadian Association of University Teachers, 1980.

26. Stillman, Paula L. *et al.* "Effect of Immediate Student Evaluations on a Multi-Instructor Course." *Journal of Medical Education* 58 (172–178): 1983.

27. Stufflebeam, D. "Meta Evaluation: An Overview." *Evaluation and the Health Professions* 1(17–43) 1978.

28. Stufflebeam, D. "A Next Step: Discussion to Consider Unifying the ERS and Joint Committee Standards" in *Standards for Evaluation Practice*, Peter H. Rossi, Editor. *New Directions for Program Evaluation* 15: 1982.

29. Thornedike, R. L. and Hagen, E. *Measurement and Evaluation in Psychology and Education*. New York: John Wiley, 1969.

30. Whitney, D. R. and Weerts, P. R. "Student Perceptions of Teaching (SPOT): I. Design for a Better System." Evaluation and Examination Service: Memo No. 25.

31. Worthen, B. R. and Sanders, J. R. *Educational Evaluation: Theory and Practice*. Belmont, California: Wadsworth, 1973.

APPENDIX A*
LIST OF POSSIBLE ITEMS FOR A TEACHING DOSSIER

THE PRODUCTS OF GOOD TEACHING

1. Scores of students on teacher-made or standardized tests, possibly "before and after."
2. Student laboratory workbooks and other kinds of workbooks or logs.
3. Student essays and project of field-work reports.
4. Publications by students on course-related work.
5. A record of students who select and succeed in advanced courses or study in the area.
6. A record of students who are able to and do elect another course with the same professor.
7. Documentary evidence of effects on student career choice.
8. Documentary evidence of help to students in securing employment.

MATERIAL FROM ONESELF

Descriptive material on current and recent teaching responsibilities and practices.

9. List of course titles and numbers, unit values or credits, enrollments, and program applicability with brief elaboration.
10. Statement about quizzes and examination items being keyed to instructional objectives.
11. List of course materials prepared for students.
12. Information on availability of students.
13. Report on identification of student difficulties and encouragement of student participation.
14. Explanation of how films or other nonprint materials were used in teaching.
15. Steps taken to emphasize the inter-relatedness and relevance of different kinds of learning.

Description of steps taken to evaluate and improve one's teaching.

16. Maintaining a record of the changes resulting from self-evaluation.
17. Reading journals on improving teaching and attempting implementation of ideas.
18. Review of new teaching materials for possible application.
19. Conducting action research on one's own teaching or course.
20. Attempts at instructional innovations.
21. Using general support services such as the Educational Resouces Information Center (ERIC) in improving one's teaching.
22. Participation in seminars, workshops and professional meetings intended to improve teaching.
23. Participation in course or curriculum development.
24. Pursuit of a line of research that contributes directly to teaching.
25. The preparation of a textbook or other instructional materials.

INFORMATION FROM OTHERS

Students

27. Student course and teaching evaluation questionnaire data which suggest improvements or produce an overall rating of effectiveness or satisfaction.

28. Written comments from a student committee to evaluate courses and provide ongoing feedback.
29. Unstructured (and possibly unsolicited) written evaluation by students, including written comments on exams and letters received after the course has been completed.
30. Documented reports of satisfaction with out-of-class contacts.
31. Interview data collected from students after completion of a course.
32. Honors received from students, such as being elected teacher of the year or most improved teacher of the year.

Colleagues

33. Statements from colleagues who have observed teaching either as members of a teaching team or as independent observers, or who teach othe sections of the same course.
34. Written comments from those who teach courses for which the present is a prerequisite.
35. Evaluation of contribution to course development and improvement.
36. Statements from colleagues from other institutions on such matters as how well students have been prepared from graduate studies.
37. Honors or recognition such as distinguished teacher award or election to a committee on teaching.
38. Requests for advice or acknowledgement thereof by a committee on teaching or similar body.

Other sources

39. Statements from administrators from one's own institution or from other institutions.
40. Alumni ratings or other graduate feedback.
41. Reports from parents of students.
42. Reports from employers of students (e.g., in a work-study or "cooperative" program).
43. Invitation to teach for outside agencies.
44. Invitation to contribute to the teaching literature.
45. Other kinds of invitations based on one's reputation as a teacher (e.g., public media, interview for a book on exemplary teaching).

*Reproduced by permission of the authors and publishers. Shore, B. M., Foster, S. F., Knapper, C. K., Nadeau, G. G., Nerill, N. and Sim. V. *Guide to the Teaching Dossier: Its Preparation and Use*. Ottowa: Canadian Association of University Teachers, 1980.

APPENDIX B
SUGGESTED FORM FOR PEER REVIEW OF UNDERGRADUATE TEACHING BASED ON DOSSIER MATERIALS

QUESTION

1. What is the quality of materials used in teaching?

DOSSIER MATERIALS

Course outline
Syllabus
Reading list
Text used
Study guide
Description of non-print materials
Hand-outs
Problem sets
Assignments

SUGGESTED FOCUS IN EXAMINING DOSSIER MATERIALS

Are these materials current?
Do they represent the best work in the field?
Are they adequate and appropriate to course goals?
Do they represent superficial or thorough coverage of course content?

Peer Reviewer's Rating: Low ___|___|___|___|___|___|___ Very High
Comments _____
_____
_____

2. What kind of intellectual tasks were set by the teacher for the students (or did the teacher succeed in getting students to set for themselves), and how did the students perform?

Copies of graded examinations
Examples of graded research papers
Examples of teacher's feedback to students on written work
Grade distribution
Descriptions of student performances, e.g., class presentation, etc.
Examples of completed assignments

What was the level of intellectual performance achieved by the students?

What kind of work was given an A? a B? a C?
Did the students learn what the department curriculum expected for this course?
How adequately do the tests or assignments represent the kinds of student performance specified in the course objectives?

Peer Reviewer's Rating: Low ___|___|___|___|___|___|___ Very High
Comments _____

3. How knowledgeable is this faculty member in subjects taught?

Evidence in teaching materials
Record of attendance at regional or national meetings
Record of colloquia or lectures given

Has the instructor kept in thoughtful contact with developments in his or her field?
Is there evidence of acquaintance with the ideas and findings of other scholars?
(This question addresses the scholoarship necessary to good teaching. It is *not* concerned with scholarly research publication.)

Peer Reviewer's Rating: Low ___|___|___|___|___|___|___ Very High
Comments _____

4. Has this faculty member assumed responsibilities related to the department's or University's teaching mission?

Record of service on department curriculum committee, honors program, advising board of teaching support service, special committees (e.g., to examine grading policies, admission standards, etc.)
Description of activities in supervising graduate students learning to teach.
Evidence of design of new courses.

Has he or she become a departmental or college citizen in regard to teaching responsibilities?
Does this faculty member recognize problems that hinder good teaching and does he or she take a responsible part in trying to solve them?
Is the involvement of the faculty member appropriate to his or her academic level? (e.g., assistant professors may sometimes become over-involved to the detriment of their scholarly and teaching activities.)

Peer Reviewer's Rating: Low ___|___|___|___|___|___|___ Very High
Comments _____

5. To what extent is this faculty member-trying to achieve excellence in teaching?

Factual statement of what activities the faculty member has engaged in to improve his or her teaching.
Examples of questionnaires used for formative purposes.
Examples of changes made on the basis of feedback.

Has he or she sought feedback about teaching quality, explored alternative teaching methods, made changes to increase student learning?
Has he or she sought aid in trying new teaching ideas?
Has he or she developed special teaching materials or participated in cooperative efforts aimed at upgrading teaching quality?

Peer Reviewer's Rating: Low ___|___|___|___|___|___|___ Very High
Comments _____

Peer Reviewer's Signature _____
Date _____

G.F. LAZOVIK 1979
UNIVERSITY OF PITTSBURGH

APPENDIX C
STUDENT EVALUATION OF LECTURE

	Strongly agree	Agree	Not sure	Disagree	Strongly disagree
1. Initially attracted my attention.	____	____	____	____	____
2. Clearly stated the purpose.	____	____	____	____	____
3. Adequately explained the relevance of the subject.	____	____	____	____	____
4. Was logical.	____	____	____	____	____
5. Kept my attention.	____	____	____	____	____
6. Highlighted key points.	____	____	____	____	____
7. Used audiovisual aids to reinforce key points.	____	____	____	____	____
8. Used conversational manner and natural gestures.	____	____	____	____	____
9. Appeared to be enthusiastic and interested.	____	____	____	____	____
10. Overall, how would you rate this lecture?	Excellent 1	2	3	4	Poor 5

Comments:

APPENDIX D
SUMMATIVE STUDENT EVALUATION OF COURSE INSTRUCTOR

1. To what degree did you like . . .	like				dislike
a. the instructor as a person?	5	4	3	2	1
b. the subject matter?	5	4	3	2	1
2. What is your rating of . . .	excellent				poor
a. lectures?	5	4	3	2	1
b. labs?	5	4	3	2	1
c. tests?	5	4	3	2	1
d. syllabus/handouts?	5	4	3	2	1
3. What is your evaluation of the job done by the instructor in teaching this course?	excellent 5	4	3	2	poor 1
4. Was there anything in particular that you liked or disliked about this course teacher?					

APPENDIX E
FORMATIVE EVALUATION RESULTS

Course: Biochemistry

Instructor: D. Jones, Ph.D.

Date of Evaluation: October 5, 1982

Lecture Topic: Disorders of Purine Metabolism

4 = Strongly Agree
3 = Agree
2 = Neutral
1 = Disagree
0 = Strongly Disagree

Number of Evaluations: 91/100 (91%)

MEAN

3.3 1. Lectured at an appropriate pace.
4 = 44 (49%) 3 = 36 (40%) 2 = 8 (9%) 1 = 1 (1%) 0 = 1 (1%)

3.7 2. Stated objectives clearly.
4 = 63 (69%) 3 = 26 (29%) 2 = 2 (2%) 1 = 0 (0%) 0 = 0 (0%)

3.8 3. Summarized major points.
4 = 73 (81%) 3 = 17 (19%) 2 = 0 (0%) 1 = 1 (1%) 0 = 0 (0%)

3.5 4. The handout was valuable in helping me learn the material.
4 = 49 (56%) 3 = 33 (38%) 2 = 6 (7%) 1 = 0 (0%) 0 = 0 (0%)

3.7 5. Overall, was effective as a lecturer.
4 = 71 (79%) 3 = 16 (18%) 2 = 2 (2%) 1 = 1 (1%) 0 = 0 (0%)

APPENDIX F
INSTRUCTIONAL VARIABLES

1. Communicates objectives and their importance to learners-answers the "so what" question.
2. Organizes material in logical sequence.
3. Presents main points clearly.
4. Uses specific examples to illustrate main point.
5. Presents material to student level of sophistication.
6. Paces appropriately for note-taking.
7. Uses appropriate instructional materials and equipment.
8. Prepares appropriate supplementary resources.
9. Cites authorities to support statements.
10. Presents divergent viewpoints for contrast and comparison.
11. Indicates own viewpoint.
12. Stimulates learner questions, responses, and discussion.
13. Summarizes to reinforce learning periodically and to achieve closure.

PERSONAL ATTRIBUTES

1. Demonstrates positive interpersonal relationships with learners.
2. Utilizes well-modulated, clearly articulated speaking voices.
3. Exhibits confidence when teaching.
4. Makes presentation without using disturbing mannerisms.

*Irby et al. (1976).

UNIVERSITY OF UTAH
HANDBOOKS FOR MEDICAL TEACHERS

1. *There Is No Gene for Good Teaching:*
   A Handbook on Lecturing for Medical Teachers
   by Neal Whitman, Ed.D. $5.00

2. *A Handbook for Group Discussion Leaders:*
   Alternatives to Lecturing Medical Students to Death
   by Neal Whitman, Ed.D and Thomas L. Schwenk, M.D. $5.00

3. *Evaluating Medical School Courses:*
   A User-Centered Handbook
   By Neal Whitman, Ed.D. and Thomas Cockayne, Ph.D. $10.00

4.* *Residents as Teachers:*
   A Guide to Educational Practice
   by Neal Whitman, Ed.D. and Thomas L. Schwenk, M.D. $5.00

*Available in Spring, 1984